CORK INSTITUTE OF TECHNOLOGY
INSTITIÚID TEICNEOLAÍOCHTA CHORCAÍ

Autumn Examinations 2017/2018

Module Title:	**Statistics and Experimental Design**

Module Code:	STAT 8004
School:	School of Science and Informatics.
Programme Title(s):	Bachelor of Engineering (Honours) in Chemical and Biopharmaceutical Engineering – Year 3 Bachelor of Science (Honours) in Environmental Science and Sustainable Technology – Award
Programmes Code(s):	**ECPEN_8_Y3** **SESST_8_Y4**
External Examiner(s):	Prof. Michael Wallace
Internal Examiner(s):	Ms. Sarah Murphy
Instructions:	Answer 3 questions. All questions carry equal marks.
Duration:	2 hours.
Sitting:	Autumn 2018

Requirements for this examination: Statistical tables by Murdoch and Barnes. Mathematical tables.

Note to Candidates: Please check the Programme Title and the Module Title to ensure that you have received the correct examination. If in doubt please contact an Invigilator.

Question 1

(a) Components of a certain type are shipped to a supplier in batches containing twenty units. Suppose that 50% of all such batches contain no defective component, 35% contain one defective component, and 15% contain two defective components. Two components are randomly selected from a batch and tested.

   i. What is the probability that neither component is defective?
   ii. What is the probability that the batch contains two defective components if neither of the selected components are defective?

(6 marks)

(b) A test contains 20 multi choice questions of which a correct answer earns a student 5 marks (100 marks in total). Each question has four possible answers, only one of which is correct. Denis has not studied for this exam and plans to guess the answer to each question.

   i. What is the probability that Denis will answer exactly half of the questions correct?
   ii. What is the probability that Denis will fail the exam (i.e. get less than 40%)?
   iii. How many questions can Denis expect to get correct?

(6 marks)

(c) The number of cars passing an automatic toll point on a motorway is modelled as a Poisson process with an average of three cars passing in a 20 second interval.

   i. What is the probability of at least three cars passing the toll in an interval of 30 seconds?
   ii. Determine the length of a time interval (in minutes) such that the probability of no car passing the toll point in this interval is 0.4.

(5 marks)

(d) Suppose that X is a continuous random variable whose probability density function is given by the following

$$f(x) = \begin{cases} \dfrac{100}{x^2} & x > 100 \\ 0 & \text{otherwise} \end{cases}$$

   i. Verify that this is a well-defined probability density function.
   ii. Find the median value of X.
   iii. What is the probability that X takes on a value less than 150?

(8 marks)

Question 2

(a) Answer each of the following questions

i. Explain what is meant by a type II error.

ii. A one sample hypothesis test was conducted using a 5% level of significance. Evidence was found to reject the null hypothesis. Will the corresponding 95% confidence interval for the true mean contain the hypothesised value or not?

iii. As a sample size increases, would you expect a standard error to increase or decrease?

(3 marks)

(b) The yield of a particular reaction has a Normal distribution with a mean of 72g and a standard deviation of 6g.

i. For a randomly chosen reaction, what is the probability that the yield lies between 68g and 77g?

ii. Twenty percent of reactions will have a yield greater than a. What is the value of a?

(6 marks)

(c) To determine the average sodium concentration of a solution, ten replicate samples were taken and the sodium content of each (mg/L) were recorded as follows:

5.655	5.392	5.500	5.486	5.749
5.548	5.765	5.653	5.654	5.841

i. Assuming normality, calculate a 98% confidence interval for the average sodium concentration of the solution.

ii. Calculate a 99% confidence interval for the variance of the sodium concentration of the solution.

(6 marks)

(d) The internal diameter of two types of washers (X and Y) are being examined. A sample of eight type X washers and seven type Y washers were taken and the internal diameter (mm) of each was recorded. These values are found in the table below

Type X	23.1	21.2	28.7	27.6	29.9	23.4	25.3	26.4
Type Y	16.5	16.3	23.9	18.4	19.6	24.0	23.4	

i. Assuming equal variances, conduct an appropriate hypothesis test to determine if there is a statistically significant difference in average interval diameter for the two types of washers. Use a 1% level of significance and clearly state your hypotheses and conclusions.

ii. Find an approximate $p-$ value for this test.

(10 marks)

Question 3

(a) A diagnostics laboratory is running tests to ensure that the method it uses to determine the fluoride level in water samples is accurate. Eight samples of water with known fluoride levels are taken and have their fluoride levels measured. The data are displayed in the table below.

Known fluoride level (ppm)	0.1	0.2	0.4	0.5	0.6	0.8	1.0	1.1
Measured fluoride level (ppm)	0.08	0.23	0.39	0.52	0.61	0.82	1.0	1.18

i. Plot the above data on a scatter diagram, and comment on the level of correlation.

ii. The equation of the least squares line is $\hat{y} = -0.012 + 1.048x$. For each of the known fluoride levels in the table above, calculate the predicted measured fluoride level and hence find the set of residuals.

iii. Using the residuals, calculate the residuals sums of squares.

iv. Produce the ANOVA table and say what conclusions you draw from this table using a 1% significance level.

v. Using relevant information from the ANOVA table, find the correlation coefficient.

(15 marks)

(b) A quality engineer at an electronics company is claiming that the distribution of defects in television sets follow a Poisson distribution. The engineer randomly selects 300 televisions and records the number of defects per television.

Defects	**Observed**
0	213
1	40
2	17
3	23
4	7

i. Is there sufficient evidence to support the engineers claim? Carry out a suitable hypothesis test using a 5% level of significance. Clearly state the hypotheses associated with the test and your conclusions.

ii. Find the $p-$ value associated with this test (as accurately as the tables allow).

(10 marks)

Question 4

(a) A manufacturer suspects that the batches of raw material delivered by the supplier differ significantly in calcium content. Of a large number of batches currently in the warehouse, three are randomly selected for use in the study. Five determinations of the calcium content are made on each batch, and the results are shown in the table below:

Batch	**Calcium Content**				
A	22.34	22.24	22.21	22.27	22.19
B	22.38	22.34	22.27	22.35	22.28
C	22.33	22.25	22.30	22.17	22.21

i. Explain what is meant by a response variable and state the response variable for this question.
ii. At a 1% level of significance, can you conclude that the average calcium content for each of the batches is the same? Clearly state the hypotheses and your conclusions.

(12 marks)

(b) An engineer is trying to improve the efficiency of a reaction that converts a raw material into a product. A 2^3experiment was designed to test the effect of three different variables; catalyst type, reaction temperature and reactant concentration on the degree of conversion. Two replicates were made at each combination of factor levels. The data are presented below:

	Temperature (B)			
	110 °		**130 °**	
	Concentration (C)			
Catalyst (A)	**0.3%**	**0.5%**	**0.3%**	**0.5%**
Type 1	59.1	45.9	54.6	22.2
	60.3	38.1	51.9	20.4
Type 2	57.6	42.3	55.2	18.6
	59.4	39.6	54.3	15.9

i. Estimate the catalyst effect, the concentration effect and the catalyst-concentration interaction effect.
ii. Test the significance of the effects estimated in part i., and state your conclusions using a 5% level of significance.
iii. Draw the catalyst-concentration interaction plot and comment.

(13 marks)

Statistical Formulae

$$Z = \frac{X - \mu}{\sigma}$$

$$\frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$$

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

$$\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$\bar{x} \pm t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}}$$

$$Z_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$(\bar{x}_1 - \bar{x}_2) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\frac{s_1^2}{s_2^2} \sim F_{n_1 - 1, n_2 - 1}$$

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2}} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$$

$$\left(\frac{(n-1)s^2}{\chi_{\frac{\alpha}{2}, n-1}^2}, \frac{(n-1)s^2}{\chi_{1-\frac{\alpha}{2}, n-1}^2}\right)$$

ANOVA

1. One-way model:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij},\ i = 1,2,...,a,\ j = 1,2,....,n.$$

Total SS: $\sum\sum y^2{}_{ij} - \frac{y_{..}^2}{an}$

Factor SS: $\sum_{i=1}^{a} \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{an}$

2. AXB factorial model:

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta_{ij}) + \varepsilon_{ijk},\ i = 1,2,...,a;\ j = 1,2,...,b;\ k = 1,2,...,n.$$

Total SS: $\sum\sum\sum y^2{}_{ijk} - \frac{y_{...}^2}{abn}$

SSA: $\sum_{i=1}^{a} \frac{y_{i..}^2}{bn} - \frac{y_{...}^2}{abn}$ SSB: $\sum_{j=1}^{b} \frac{y_{.j.}^2}{an} - \frac{y_{...}^2}{abn}$

3. 2^k design, n replicates.

Effect estimate given by $\frac{(Contrast)}{n.2^{k-1}}$

Effect SS given by $\frac{(Contrast)^2}{n.2^k}$

$V(\text{effect estimate}) = \frac{{\sigma_e}^2}{n.2^{k-2}}$, where ${\sigma_e}^2$ is the error variance